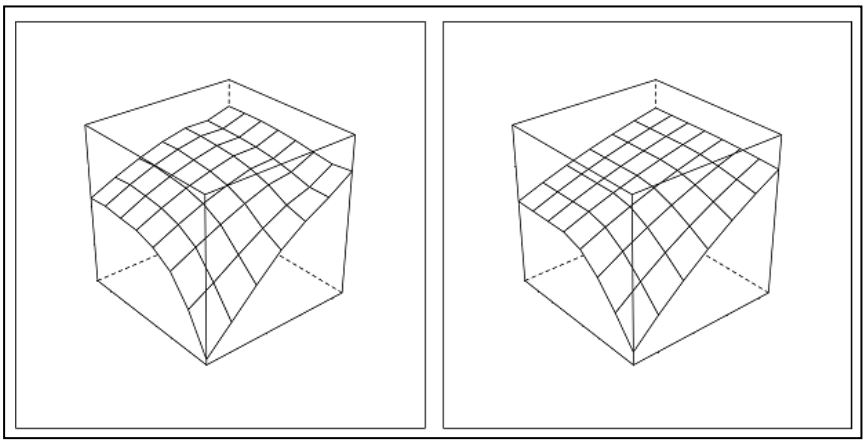
Spatio-temporal analysis and interpolation of PM_{10} measurements in Europe

Including Erratum

(dd. 18 March 2013)



ETC/ACM Technical Paper 2011/10

November 2011

Benedikt Gräler, Lydia Gerharz, Edzer Pebesma

The European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM)
is a consortium of European institutes under contract of the European Environment Agency
RIVM UBA-V ÖKO AEAT EMISIA CHMI NILU INERIS PBL CSIC

Front page picture:
One of the tested spatio-temporal dependence models in this paper, Figure 3 with empirical (left) and fitted (right) spatio-temporal variogram of the residuals of log-transformed PM_{10} concentrations used in the product-sum covariance method.

Author affiliation:
Benedikt Gräler, Lydia Gerharz, Edzer Pebesma: Institute for Geoinformatics (IfGI), University of Muenster, Germany.
(Under subcontract of ETC/ACM Consortium partner institute RIVM, NL).

DISCLAIMER

ETC/ACM Technical Paper 2011/10
European Topic Centre on Air Pollution and Climate Change Mitigation
PO Box 1
3720 BA Bilthoven
The Netherlands
Phone +31 30 2748562
Fax +31 30 2744433
Email etcacm@rivm.nl
Website http://acm.eionet.europa.eu/

INSTITUTE FOR GEOINFORMATICS, UNIVERSITY OF MUENSTER

Spatio-temporal analysis and interpolation of PM_{10} measurements in Europe

Task 1012 - Subtask 4

ETC/ACM Technical Paper 2011/10

Benedikt Gräler, Lydia Gerharz, Edzer Pebesma
ben.graeler@uni-muenster.de
November 2011

Including Erratum
(dd. 18 March 2013)

Erratum

The study presented in the ETC/ACM Technical Paper 2011/10 investigated the potential of spatio-temporal kriging approaches for daily mean PM_{10} concentrations. The methods used include separate daily variogram estimates, temporally evolving variograms, the metric model, the separable covariance model and the product-sum model, and are combined with multiple linear regression. These methods are applied to daily mean rural background PM_{10} concentrations across Europe for the year 2005, and incorporate daily EMEP model data and elevation data as predictors.

Unfortunately, the EMEP model data that has been incorporated in some of the methods has been distorted in the preparation for the analysis. This flaw has now been detected and corrected. An updated table summarizing all new results (as Table 3 in the original report) is presented below. The linear model underlying the residual kriging approaches now has a considerable higher explanatory value (an increase of the adjusted R^2 value from approx. 10% to 30%). In general, the cross-validation statistics are of the same order of magnitude as in the published report. However, the strength of the spatio-temporal approaches on a daily cross-validation is weakened due to the better regression. Nevertheless, the statistics on the predicted number of days exceeding the threshold of 50 µg/m³ improve. The best approaches on a yearly level remain the metric kriging approaches using 100 and 1000 neighbours.

Replacement for Table 3: Statistics for the cross-validation of the methods with the corrected EMEP data. Yellow cells score second best, green cells score best

	direct kriging			multi linear regression		multi linear regression + residual Kriging				log multi linear regression + residual Kriging											
2005	yrl. means	yrl. log means	3D [100]	once	log, once	TP 2007/8 (3a)	3D [100]	sep. days	mean	yrl. means	sep. days	evo: l=0.9	evo: l=0.5	evo: l=0.1	poole d	mean	mow	3D [100]	3D [1000]	sep. COV	p-s. COV
ID	none	none	1-e	2	3	none	2-e	2-a	2-d	none	3-a	3-b	3-b	3-b	3-c	3-d	3-e	3-f	3-f	3-g	3-h
Daily																					
RMSE			10.69	13.62	14.38	9.77	10.34	10.29	9.91		10.29	10.28	10.28	10.34	10.35	10.26	10.27	10.67	10.66	10.14	10.26
BIAS			-0.29	0.00	2.76	-0.04	-0.20	-0.32	-0.11		-0.32	-0.31	-0.29	-0.33	-0.34	-0.50	-0.32	-0.47	-0.35	-0.11	-0.44
MAE			6.28	8.68	8.52	5.67	6.14	5.80	5.67		5.80	5.79	5.79	5.83	5.84	5.81	5.79	6.16	6.16	5.96	5.91
adj. R^2			0.53																		
COR			0.73	0.48	0.44	0.78	0.75	0.75	0.77		0.75	0.76	0.75	0.75	0.75	0.76	0.76	0.74	0.74	0.74	0.75
Yearly																					
RMSE	5.84	5.88	5.96	6.60	7.22	4.77	4.93	4.87	4.95	4.82	4.87	4.88	4.92	4.97	4.97	4.90	4.92	4.66	4.57	5.44	5.26
BIAS	-0.13	-0.06	-0.28	0.01	2.76	-0.02	-0.22	-0.34	-0.12	0.20	-0.34	-0.33	-0.31	-0.35	-0.36	-0.52	-0.34	-0.49	-0.37	-0.46	-0.10
MAE	4.09	4.10	4.13	4.92	4.87	3.32	3.43	3.42	3.45	3.40	3.42	3.43	3.45	3.50	3.52	3.45	3.45	3.29	3.28	3.99	3.81
adj. R^2	0.38	0.37	0.37																		
COR	0.62	0.61	0.61	0.46	0.45	0.77	0.75	0.75	0.74	0.77	0.75	0.75	0.75	0.74	0.74	0.75	0.75	0.78	0.79	0.68	0.72
NOE																					
RMSE			19.09	25.71	24.95	16.78	16.47	15.69	16.96		15.69	15.71	15.84	16.02	16.14	15.49	15.87	14.36	13.84	19.97	18.43
BIAS			4.84	13.02	11.44	4.43	4.87	2.88	4.70		2.88	2.93	2.93	2.88	3.19	2.44	2.88	2.19	2.42	8.06	7.17
MAE			10.03	13.54	12.85	8.63	8.74	8.56	8.84		8.56	8.53	8.60	8.68	8.78	8.58	8.63	7.72	7.30	10.69	10.40

Content

Summary

This study investigates the potential of spatio-temporal kriging approaches for daily mean PM_{10} concentrations. The methods used include separate daily variogram estimates, temporally evolving variograms, the metric model, the separable covariance model and the product-sum model, and are combined with multiple linear regression. These methods are applied to daily mean rural background PM_{10} concentrations across Europe for the year 2005, and incorporate daily EMEP model data and elevation data as predictors.

The air quality indicators used in this study are the daily and yearly mean PM_{10} concentrations and the number of days exceeding the limit value 50 µg/m³ (NOE). The quality assessment of the different techniques relies on a cross-validation. Statistical measures are used to quantify the improvement for different indicators.

It is shown that daily interpolations can improve the statistical performance of the interpolation of annual mean PM_{10} concentrations, by a 7% reduction of RMSE in cross-validation. Furthermore, some advantages of daily estimates are described. Besides the improvement in annual mean PM_{10} concentration maps, studying the phenomenon in a wider spatio-temporal context becomes possible with daily estimates. Especially the estimation the number of days PM_{10} concentrations exceed certain limits can be done in a more natural way. Likewise, the detection of outliers and data inhomogeneity benefits from a daily spatio-temporal model.

Interpolation with the simple spatio-temporal variogram models used here exploits the temporal correlations present and out performs the purely spatial interpolation methods by a 5-6% reduction of RMSE in cross-validation. Based on temporal variability of the spatial short-distance variation component, a discussion is given on the suitability of this statistic to infer measurement errors, and alternative approaches are proposed.

As in this study the methods used for modelling spatio-temporal dependencies are very simple, we recommend to investigate the value of more comprehensive methods for this.

1 Introduction

The increasing number of sensors and measurement stations with a finer temporal resolution call for sophisticated spatio-temporal interpolation methods. Investigating a series of maps showing the daily mean concentration of PM_{10} reveals significant temporal and hence spatio-temporal dependencies. A purely spatial interpolation approach ignores these dependencies and can be seen as a spatio-temporal interpolation where all temporal correlations are set to zero. Obviously, investigating this additional source of information has some potential to improve the mapping of PM_{10} and spatio-temporal processes in general. Besides a few simple approaches, this report evaluates three full spatio-temporal kriging techniques.

1.1 Aim and outline of the study

This study was conducted to further explore the potential of different spatio-temporal kriging methods for daily PM_{10} measurements across Europe. Therefore, a set of different approaches has been implemented in R (R Development Core Team, 2011). Daily means from AirBase (EEA: AirBase version 5) across Europe for the year 2005 serve as a sample data set. A cross-validation study has been conducted and a set of different quality measures has been derived for every method. The cross-validations have been exploited at two temporal resolutions:

- daily, comparing the predicted daily mean concentrations, and
- yearly, looking at the station-wise annual mean of daily predictions.

Additionally, the cross-validations compare the number of exceedance days (NOE) on a yearly level

- NOE, i.e. the number of days where the predicted daily mean exceeds 50 µg/m³ in 2005 have been compared to the observed number of daily exceedances.

1.2 Spatio-temporal interpolation methods

Several spatio-temporal interpolation methods have been applied. The approaches are set-up similarly to the interpolation methods presented in Denby et al. (2008a), to allow for an easy comparison between the ever growing set of new methods. Furthermore, annual interpolation methods, similar to those currently operationally implemented by ETC/ACM, are also considered for comparison.

The set of evaluated spatio-temporal kriging methods is applied to differently pre-processed daily mean PM_{10} concentrations as provided by AirBase. We consider:

1. the unaltered daily means
2. residuals after multiple linear regression of the observed daily mean concentrations with altitude, EMEP model data and a periodic temporal predictor
3. residuals after multiple linear regression of the logs of observed daily mean concentrations with altitude, EMEP model data and a periodic temporal predictor

The kriging methods we investigate include:

a. daily estimation of the variogram (see 2.2)
b. evolving variograms using previous variograms to weight the current one (see 2.3)
c. the constant pooled variogram over the year (see 2.4)
d. the constant mean variogram over the year (see 2.4)
e. moving window variogram estimation (see 2.5)

The aforementioned methods only make use of the temporal information through the variogram estimation. In additional three fully spatio-temporal methods were applied:

f. a metric model (3D kriging, see 2.6)
g. a separable covariance function (see 2.7)
h. a product-sum covariance function (see 2.8)

The separable and product-sum covariance functions assume the spatial and temporal correlation processes to be disjoint for space and time. The 3D kriging assumes time to be a third orthogonal dimension with identical correlation structure for spatial, temporal and spatio-temporal distances. Therefore, a rescaling of the temporal domain is necessary to reduce anisotropy in 3D kriging.

Finally, the regression predictions and their residuals are united and back-transformed where necessary. The desired yearly averages are calculated per interpolated grid cell throughout the year.

1.3 Assessment methodology

The quality of all methods is assessed using statistical measures and applicability concerns in terms of processing time. The comparison of different interpolation methods takes place on a yearly (comparing yearly means for predictions and measurements) and daily (comparing daily means for predictions and measurements) level as well as in terms of the NOE. The root mean squared error (RMSE), the bias (BIAS), mean absolute error (MAE) and Pearson's correlation coefficient (COR) between predictions and measurements at every location are calculated using leave-one-out cross-validation. Furthermore, a linear model for predictions and measurements is fitted and the adjusted R^2 considered as a fifth measure to assess the prediction quality. All methods are compared using the above measures where applicable. As reference values of the statistical measures, we provide values for different approaches relying on simpler models as well. All methods are implemented in R (R Development Core Team, 2011) and heavily rely on the R packages *spacetime* (Pebesma, 2011) and *gstat* (Pebesma, 2004).

1.4 Data

The daily mean PM_{10} concentrations test data is obtained from EEA's AirBase (version 5) for all European countries. To enhance comparability to former interpolation approaches in Denby et al. (2008a), all methods are applied to daily mean values in 2005 of rural background stations. Different than in the yearly interpolation approach in Horálek et al. (2007), we used all measurement stations and not only those providing a data integrity of at least 75% throughout the year. Thus, the number of measurement stations amounts to 194. However, when considering the complete space-time lattice, i.e. where every space and time combination is available, the data set yields 2568 missing values (3.6%). Some methods benefit from the spatio-temporal setting and can correct for these missing values while others require a complete spatio-temporal data set. Details will be given in the following section.

Daily mean PM_{10} concentrations derived from model data by the European Monitoring and Evaluation Programme (EMEP) serve as additional data in the multiple linear regression. The 50 km EMEP Polar Stereographic grid is converted and projected to match the 10 km gridded interpolation domain in the standard EEA ETRS89-LAEA5210 projection. Additionally, altitude data is used from both the AirBase measurement stations and from the EEA 10 km grid originating from the 30 arcsec GTOPO altitude grid (WGS84) as applied in Horálek et al. (2007).

2 Details on spatio-temporal interpolation methods

We consider a range of different spatio-temporal kriging methods varying from simple daily interpolations up to a full 3D approach, each highlighted in this chapter.

The interpolation methods consist of using:

a. separate daily variograms (see 2.2)
b. daily evolving variograms with different weighting for current and previous day's variogram (see 2.3)
c. pooled variograms (see 2.4)
d. mean variograms (see 2.4)
e. moving window variograms (see 2.5)
f. complete 3D variogram with either 100 or 1000 nearest neighbours (see 2.6)
g. separate covariance function (see 2.7)
h. product-sum covariance function (see 2.8

The underlying data are

1. daily mean PM_{10} concentrations,
2. residuals after a multiple linear regression of the concentrations, and
3. residuals after multiple linear regression of the log-transformed concentrations.

The spherical variogram model repeatedly appears to be best suited to capture the spatial and temporal dependencies and is therefore adopted in all methods. Its parameters are fitted using the functions provided by the R package *gstat* (Pebesma, 2004). The R scripts that were used are available from the authors upon request.

To support the multi-linear regressions the PM_{10} concentrations and their log-transforms undergo some pre-processing (section 2.1). When adding the interpolated residuals again to the estimates produced by the multi linear regression some post-processing (section 2.9) is needed to compensate for bias in case of using log-transformed data.

2.1 Pre-processing

A multiple linear regression is applied to the daily mean PM_{10} concentrations from the AirBase database and their log-transforms. The predictors include:

- the daily EMEP grid cell's prediction value for every station
- a temporal index to adjust for periodic yearly changes: $\cos(\frac{\text{day}}{365} \cdot 2\pi)$
- station altitude

In contrast to the yearly interpolation approach for rural areas (Horálek et al., 2007), meteorological data (wind speed, solar radiation) are not used here as the predictors.

Variations in the combination of the chosen predictors are also tested. This set-up is used for multiple linear regression of the daily mean and the log transformed daily mean PM_{10} concentrations. An overview of the estimated regression coefficients for the different parameters can be found in Table 1 below.

Table 1: Multiple linear regression coefficient estimates and standard errors for daily mean PM_{10} concentrations and their log-transforms. The adjusted R^2 is given in square brackets.

multi linear regression coefficients …				
	… for PM_{10} [0.05]		… for log(PM_{10}) [0.10]	
	Estimate	Std. Error	Estimate	Std. Error
Intercept	23.7929	0.0996	3.0190	0.0042
EMEP model	0.0800	0.0091	0.0035	0.0004
Time	3.5697	0.1392	0.1195	0.0058
altitude	-0.0089	0.0002	-0.0006	0.0000
model:time	-0.0848	0.0126	-0.0040	0.0005
time:altitude	-0.0053	0.0002	-0.0004	0.0000

Figure 1 shows the normal quantile-quantile-plots for both linear regression models. These plots are well suited to visually judge the underlying assumption of normal distributed residuals. Perfectly normal distributed residuals would follow the dotted diagonal line.

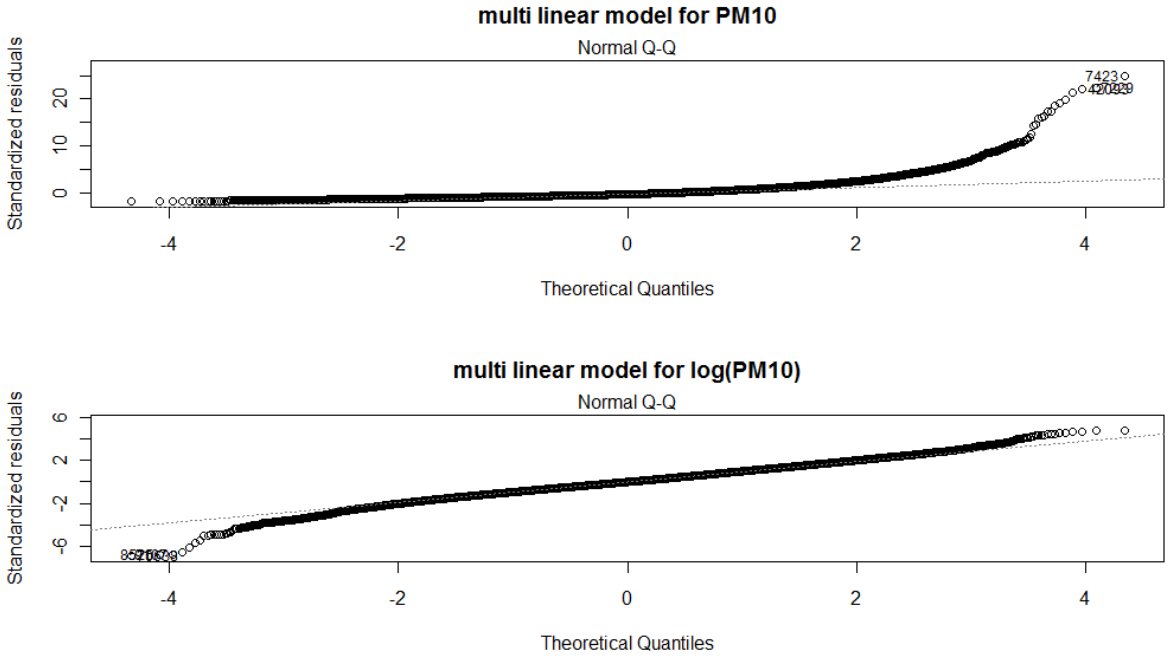


Figure 1: Diagnostic plots showing the normal quantile-quantile-plots for both regression models.

Referring to the upper plot, the assumption of normal distributed residuals does not hold as the dots heavily deviate from the line on the right side. This indicates a right skewed data set. However, the residuals of the log-transformed data do seem to fulfil this requirement. Therefore and due to the higher adjusted R^2 value, most of the results presented in this report refer to the log-transformed multiple linear regression. For comparison, a few results are based on the non-transformed regression as well.

2.2 Separate daily variograms (a)

The simplest approach is to interpolate the daily mean PM_{10} concentrations for every single day separately by independently fitted daily variograms. This allows the spatial dependence pattern to change on a daily basis. However, large local variability for single days may result in singular variogram estimates neglecting any spatial dependence. Kriging is then performed

for every time slice separately. The temporal component of this approach is reflected in its flexibility allowing for changing spatial dependence structures. This might be advantageous compared to the interpolation of yearly mean PM_{10} concentrations with a single spatial variogram. This is the kriging method applied by Denby et al. (2008a).

2.3 Daily evolving variograms (b)

Assuming temporal continuity in the spatial dependence patterns that might be partially covered by white noise processes raises the idea of daily evolving variograms. The estimation of the variogram's parameters is to a certain degree controlled by the history. In order to allow for a flexibly varying sill for every day, it is estimated from the current day only. The $nugget/sill$ ratio is then controlled using previous estimates. This method introduces an additional parameter that controls the weighting between current and previous estimates as follows:

$$\begin{aligned} range &= \lambda \cdot range_{curr} + (1-\lambda) \cdot range_{prev} \\ nugget &= \left(\lambda \cdot \frac{nugget_{curr}}{sill_{curr}} + (1-\lambda) \cdot \frac{nugget_{prev}}{sill_{prev}}\right) \cdot sill_{curr} \\ partial\ sill &= sill_{curr} - nugget \end{aligned}$$

Where the subscript $curr$ indicates the estimate of the current day and $prev$ refers to the weighted estimate of the previous day. The recursive structure includes all preceding days but with an exponentially decaying influence. As before, kriging is performed time slice wise, but the estimation of the variograms to some degree utilizes the temporal information.

2.4 Pooled variogram (c) and mean variogram (d)

Assuming a constant spatial structure throughout the year leads to another extreme case where every day is interpolated using the same variogram. Averaging the empirical variograms for every lag class leads to the *pooled variogram*. Every day is treated as a copy of the same spatial dependence structure by taking the mean of all daily semi-variances within each lag-class. An alternative approach consists of averaging all variogram parameters (i.e. nugget, partial sill and range) of the non-singular daily estimates referred to as *mean variogram*. The estimated parameters of the spherical variograms of both approaches are given in Table 2 for comparison.

Table 2: Estimated variogram parameters for different spherical variogram models for the residuals of the log-transformed PM_{10} concentrations.

	pooled variogram (c)	mean variogram (d)
Nugget	0.0916	0.0842
partial sill	0.2538	0.2477
range [m]	880534	568081

2.5 Moving window variograms (e)

This method is a union of using separate daily and constant variograms. The variograms are calculated as the mean variogram for a moving window. Beside the current estimate, estimates of three preceding and three following days are used to average the variogram parameters. This method allows for daily changing variograms without only relying on the structure present in a single day. Kriging is again performed for every day separately utilizing the temporal information in the estimation process only. Adaption of seasonal or trend components in the spatio-temporal dependence structure is the strength of this method.

2.6 Time as a 3rd dimension (f)

Considering time to be the third orthogonal dimension leads to a natural 3-dimensional extension of 2-dimensional kriging. In order to allow for meaningful results, the temporal dimension has to be rescaled to align with the spatial directions. A reasonable fit (by visual comparison of the spatial and temporal variograms) could be achieved assuming one day to correspond to a distance of 120 km, which reflects the ratio of the ranges of the spatial variogram and the temporal variogram. With this rescaling, the set of the most correlated neighbours in space and time will have the strongest influence on the prediction. Otherwise, mostly temporal neighbours would be used showing no more dependence after approximately 6 days. This method relies on a single variogram (nugget: 0.09, partial sill: 0.29, range: 816859 m; see Figure 2) and predictions can be made for any point in the space-time cube. Therefore, this approach enables us to generate transition maps between days. The interpolation is no longer restricted to single time slices and every location is estimated based on its 100 nearest neighbours in the space-time cube (note that predictions made for early January mostly depend on following observations where predictions in late December mostly rely on past observations). Additionally, 1000 neighbours were selected to study the influence of the neighbourhood. In this approach, missing values in a single time slice are made up for by using values from the preceding and following time slices. Thus, all the available data across space and time is incorporated.

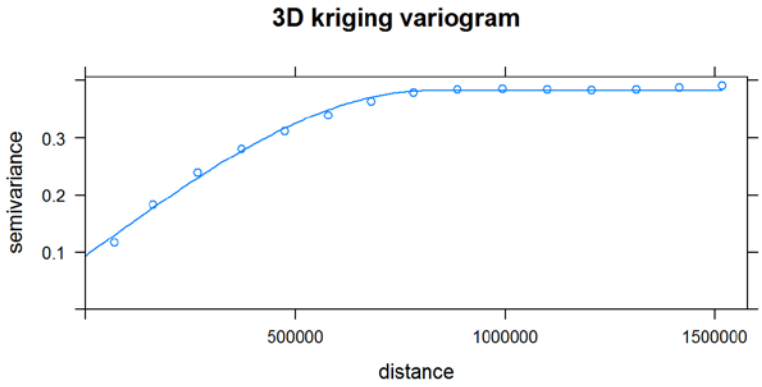


Figure 2: Variogram of the 3D kriging approach of the residuals of the log-transformed PM_{10} concentrations. Distance corresponds to meters, where a temporal difference of 1 day equals 120 km.

2.7 Separable covariance function (g)

Extending the kriging approach from space to space-time alters the covariance function. In space and under the assumptions of isotropy and stationarity, the covariance is a function $C(h)$ of the separating distance h between two locations. A spatio-temporal covariance function is thought of as a function of a spatial and a temporal distance $C(h,t)$. In general, this function is hard to estimate. A separable covariance function is assumed to fulfil $C(h,t) = C_S(h) \cdot C_T(t)$ and the distinct spatial and temporal effects can be estimated with relative ease. The interpolation is performed based on the complete spatio-temporal data set provided for the variogram estimation. This allows predicting values at any location and point in time.

A drawback of this approach potentially affecting the predictions occurs when one has to deal with incomplete data sets. The estimation of the kriging weights relies on identical and complete time series for every spatial location. In order to adopt for a large number of missing values in the AirBase data set, a month wise spatio-temporal interpolation is performed. However, the spatial and temporal variograms used in the monthly interpolations are fitted once for the full set of observations only. Thus, the predictions are based on a discrete set of

temporal neighbourhoods. The spatial variogram is the mean variogram as modelled before, and for the temporal variogram we use the spherical model and all available time series were taken as representative for the same temporal process (nugget: 0.10, partial sill: 0.48, range: 5.5 days).

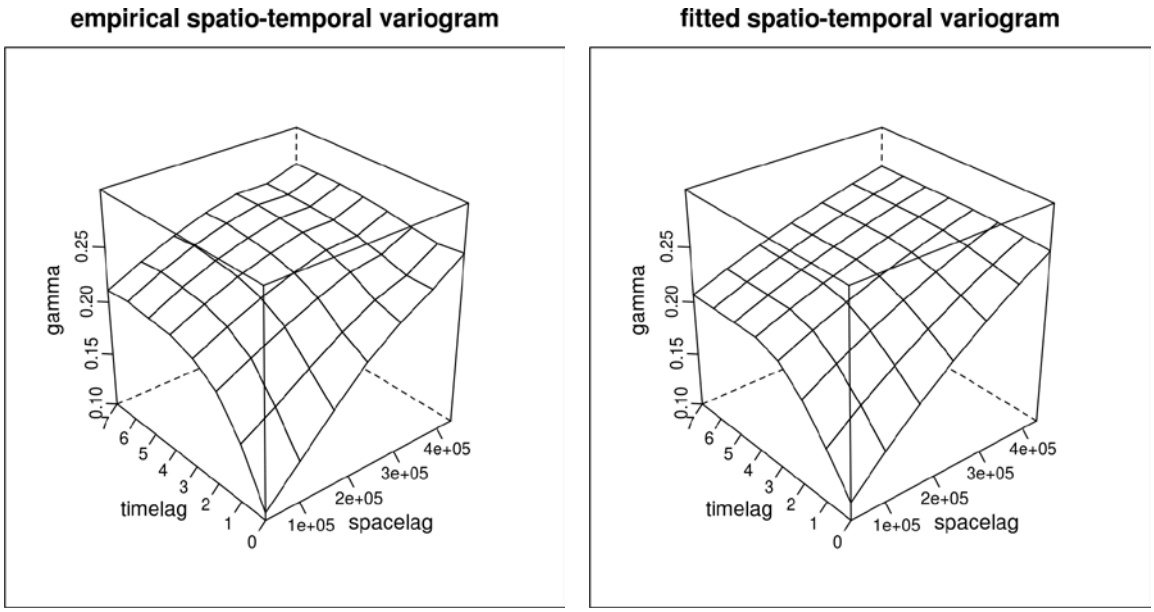


Figure 3: Empirical and fitted spatio-temporal variogram (gamma) of the residuals of log-transformed PM_{10} concentrations used in the product-sum covariance method.

2.8 Product-sum covariance function (h)

A powerful extension to the above is the introduction of the product-sum covariance function. It relies on the assumption that the spatio-temporal covariance function $C(h,t)$ can be written as $C(h,t) = k_1 C_S(h) \cdot C_T(t) + k_2 C_S(h) + k_3 C_T(t)$ where $k_1 > 0$ and $k_2, k_3 \geq 0$ are real coefficients. We follow an estimation approach introduced by De Iaco (2001) where the coefficients k_1, k_2 and k_3 depend on the spatial, temporal and spatio-temporal sills. Figure 3 shows the empirical and the estimated spatio-temporal variogram. The extended covariance model gains flexibility in modelling the spatio-temporal process. This approach has a similar technical drawback as the one before: the estimation can only be based on complete spatio-temporal datasets. Thus, we follow the same month-wise interpolation work around using disjoint temporal neighbourhoods which might as well affect the predictions. Additionally, we restrict the method to all locations within a range of 1200 km, otherwise the necessary inversion of large matrices in the kriging process would fail due to the very small weights of far apart stations.

2.9 Post-processing

After the residuals have been interpolated, they are added again to the estimates produced by the multiple linear regression. In general, the kriging predictor is asymptotically unbiased and will give the correct estimate. In our case, the estimates are correct for the log-transformed

concentrations. The non-linear back-transformation using exp alters the weights of the linear prediction on the non-log scale resulting in biased predictions. Unfortunately, a full correction for this bias of the joint linear regression and kriging predictor cannot be derived and we are restricted to use the kriging variance $\sigma^2(s)$ of the residuals at each prediction location s as an estimate. Predictions that have been log-transformed in the pre-processing step are back-transformed using $Y(s) = \exp(Z(s) + \sigma(s)^2/2)$ including the correction factor to lower the effect of the unbiased estimator (Denby et al., 2008b). Here, $Z(s)$ denotes the estimate at some location s combining the multiple linear regression model with the kriged residuals. The kriging variances are back-transformed by $V(s) \coloneqq (\exp(\sigma^2(s)) - 1) \cdot \exp(2\,Y(s) + \sigma^2)$ (Equation 2, Denby et al., 2008b). The annual PM_{10} averages are given by the means per grid cell of all daily interpolated values and the annual variance is roughly approximated by the formula given by Denby et al. (2008b, Appendix A). The number of exceedance days is calculated by checking for exceedances at a daily prediction level.

3 Results

In general, most of the spatio-temporal interpolation techniques perform reasonably well. None of the cross-validation scatterplots shows a completely different structure than the others. However, a method in favour can be selected based on the statistical analysis.

3.1 Statistical measures

The sequential cross-validation leaving each station out once and predicting the measurements from the remaining ones is the key to compare the different interpolation approaches. For every cross-validation a set of statistical measures is calculated: the root mean square error (RMSE) providing the standard deviation of the errors, the bias (BIAS) and mean absolute error (MAE) indicating the mean and mean absolute deviation of the prediction. Furthermore, Pearson's correlation coefficient and the adjusted R^2 of a linear model for predictions vs. measurements are calculated. The full set of all cross validation results is given in Table 3 which compares the listed methods on a daily and yearly basis. The row denoted with 'ID' refers to numbers and letters assigned to the methods introduced in Section 1.2. Table 3 provides the set of benchmarks for:

- direct kriging
- multiple linear regression only
- results for different kriging approaches using the residuals after multiple linear regression of PM_{10} concentrations
- results for different kriging approaches using residuals of the multiple linear regression of log-transformed PM_{10} concentrations.

Each interpolation approach is a composition of the pre-processing strategy (indicated by 1 – 3) and the kriging method (indicated by a – h), and indicated in the blue header cells. Thus, some kriging techniques appear several times in Table 3, but under different pre-processing strategies. Every column defines a unique interpolation approach.

The three direct kriging methods (first three columns) do not use residuals of a multi linear regression and are applied directly to the measurements. The direct methods "yearly means" and "yearly log means" interpolate the measurement stations' yearly (log) mean values across Europe. Hence, they cannot be compared on a daily basis and do not predict days exceeding the limit concentration and are thus mainly given as reference values.

The second set of columns entitled "multi linear regression" describes pure regression models predicting the PM_{10} concentrations in Europe. The residuals of these approaches provide ultimately the data for our kriging approaches of which their statistical results are presented in the third and fourth set of methods. The multiple linear regression used at the method "TP 2007/8 (3a)" differs from all regressions introduced earlier in this report and is therefore not represented in the table at this second set of methods. The approach "TP 2007/8 (3a)" uses residuals from daily multiple linear regression models of the EMEP model data and altitude.

We re-implemented the method described by Denby et al. (2008a) and re-performed their analysis with all available data as in our approaches to increase comparability. The third and fourth set of methods represents their statistical results.

Table 3: Statistics for the cross-validation of the benchmarking of the methods. Yellow cells score second best, green cells score best. The red numbers in the NOE rows indicate the worst under-prediction of 0 exceedance days. (NA = not applicable).

Erroneous values in this table. See Erratum at front of paper with corrected values in the replacement table.

	direct kriging			multi lin. regression		multi lin. regression + residual Kriging				log multi linear regression + residual Kriging											
2005	yrl. means	yrl. log means	3D [100]	once	log, once	TP 2007/8 (3a)	3D [100]	sep. days	mean	yrl. means	sep. days	evo: λ=0.9	evo: λ=0.5	evo: λ=0.1	pooled	mean	mow	3D [100]	3D [1000]	sep. COV	p-s. COV
ID	none	none	1-e	2	3	none	2-e	2-a	2-d	none	3-a	3-b	3-b	3-b	3-c	3-d	3-e	3-f	3-f	3-g	3-h
Daily																					
RMSE	NA	NA	10.69	15.16	15.68	10.04	10.10	9.86	9.80	NA	9.86	9.85	9.87	9.91	9.92	9.87	9.84	9.84	9.86	9.63	9.37
BIAS	NA	NA	-0.29	0.00	3.70	-0.07	-0.25	-0.14	-0.15	NA	-0.14	-0.14	-0.14	-0.21	-0.12	-0.36	-0.15	-0.24	-0.21	0.59	0.49
MAE	NA	NA	6.28	10.17	9.68	5.91	5.84	5.65	5.61	NA	5.57	5.57	5.57	5.60	5.60	5.59	5.56	5.66	5.71	5.61	5.50
adj. R^2	NA	NA	0.53	0.05	0.04	0.58	0.58	0.60	0.60	NA	0.60	0.60	0.60	0.59	0.59	0.60	0.61	0.60	0.60	0.59	0.61
COR	NA	NA	0.73	0.22	0.20	0.76	0.76	0.77	0.77	NA	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.78
Yearly																					
RMSE	5.84	5.88	5.96	6.77	7.78	4.87	4.97	4.96	4.98	5.04	4.93	4.94	4.98	5.01	5.03	4.98	4.97	4.66	4.76	5.31	5.14
BIAS	-0.13	-0.06	-0.28	0.03	3.72	-0.05	-0.24	-0.12	-0.14	-0.20	-0.12	-0.12	-0.12	-0.19	-0.10	-0.34	-0.13	-0.22	-0.19	0.34	0.19
MAE	4.09	4.10	4.13	4.95	5.25	3.35	3.46	3.40	3.43	3.49	3.31	3.32	3.34	3.38	3.41	3.35	3.33	3.24	3.34	3.71	3.67
adj. R^2	0.38	0.37	0.37	0.16	0.15	0.58	0.55	0.55	0.55	0.54	0.55	0.55	0.54	0.53	0.53	0.54	0.58	0.61	0.59	0.49	0.52
COR	0.62	0.61	0.61	0.41	0.40	0.76	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.73	0.73	0.74	0.74	0.78	0.77	0.70	0.72
NOE																					
RMSE	NA	NA	19.09	27.30	27.30	17.23	17.37	17.39	17.54	NA	16.25	16.26	16.44	16.44	17.10	16.09	16.56	14.85	15.18	20.62	20.91
BIAS	NA	NA	4.84	15.43	15.43	4.84	4.85	5.06	5.08	NA	4.12	4.11	3.99	3.75	4.66	3.38	4.04	3.95	4.10	10.12	9.98
MAE	NA	NA	10.03	15.43	15.43	8.66	9.08	8.86	8.90	NA	8.21	8.28	8.39	8.36	8.72	8.39	8.41	7.88	7.94	10.84	11.10

Comparing the results on a yearly level, the two 3D kriging approaches using residuals of a multiple linear regression on log-transformed measurements (with 100 and 1000 neighbours in space and time) perform best. Especially the prediction of Number of Exceedance days (NOE) shows a clear gain. Even if both methods have a considerable bias, they seem to capture the exceedance structure better than the competitive approaches. It is interesting to notice that the approach relying on fewer neighbours performs better.

The ranking on a daily level is not as straight forward as on a yearly level. In terms of adj. R^2 and COR, most approaches are hard to distinguish from each other. The lowest bias was found for the method following the one by Denby et al. (2008a), "TP 2007/08 (3a)". However, RMSE and MAE are within the range of the larger ones. Using the spatio-temporal product-sum covariance model, "p-s. COV (3-g)", results in the smallest RMSE and MAE. Unfortunately, its bias is one of the highest.

It is interesting to notice that the multiple linear regression presented in this report performs worse than regressions presented earlier (Denby et al., 2008a) although the overall prediction improves. Thus, the residuals seem to provide a stronger spatial dependence structure.

Two cross-validation scatter plots on a daily and yearly level are given in Figure 4. They are based on the multiple linear regression of the log-transformed PM_{10} daily mean values and 3D kriging over 100 neighbours (see Section 2.6). The plots show a reasonable fit on the yearly level where the daily cross-validations still have pretty large deviations. Extreme outliers observed in the daily comparison are discussed in more detail in chapter 4.

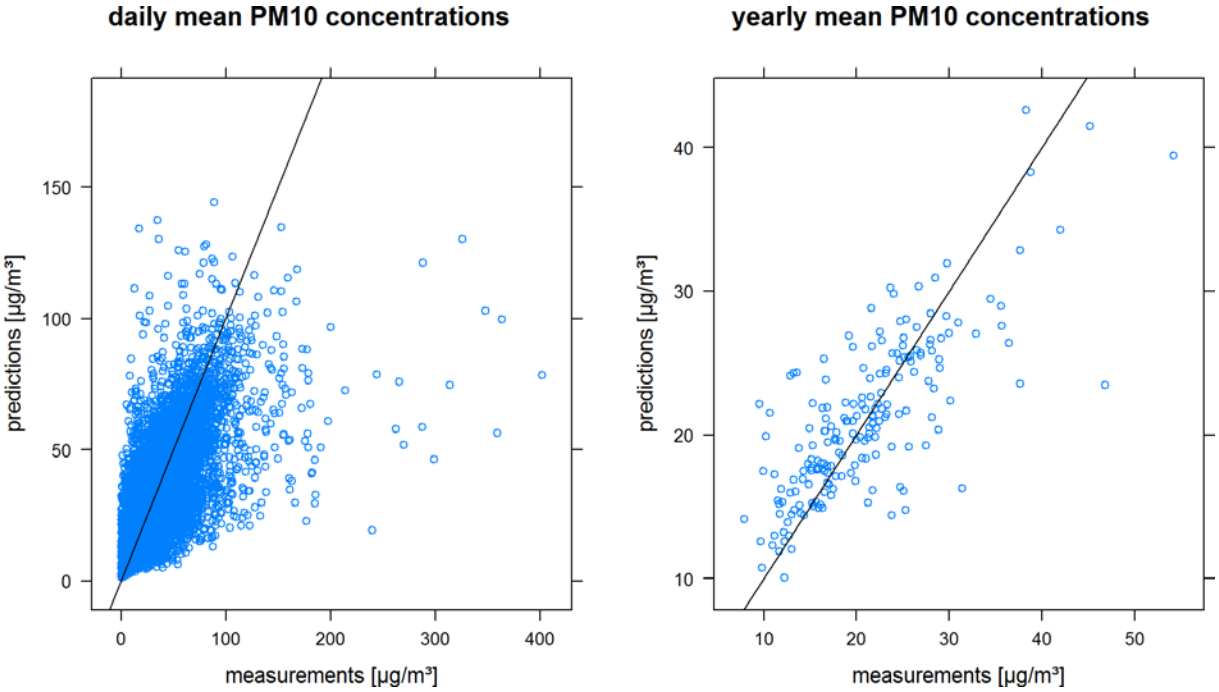


Figure 4: Two cross-validation plots of the best performing method for yearly and daily mean PM_{10} concentrations.

3.2 Maps

We chose the best performing procedure according to the statistical measures on a yearly level (3D with 100 neighbours, 3-f) to produce a map of rural background concentrations over Europe in Figure 5. The standard deviation shown in Figure 5 (right) approximates the standard deviation of the predicted annual means. It is derived as described by Denby et al. (2008b, Appendix A) using a temporal correlation coefficient of 0.3. However, block kriging (Cressie and Wikle, 2011) where the blocks are years would provide more correct values than the approximate approach suggested by Denby et al. (2008b).

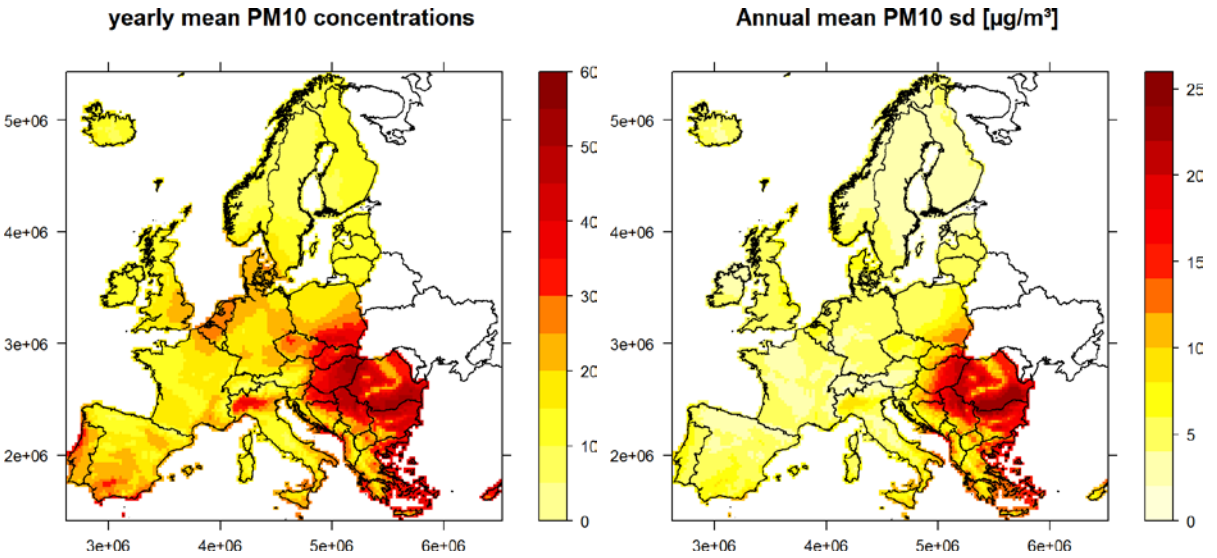


Figure 5: Interpolated map of yearly mean PM_{10} concentrations in Europe 2005 (left) and the approximate annual standard deviation of the predictions calculated as by Denby et al. (2008b) using a temporal correlation coefficient of 0.3 (right). The method adopted is the 3D interpolation using 100 neighbours.

Figure 6 shows the observed and estimated number of exceedance days. Some of the very high values in Eastern Europe have to be read with care. The high kriging variances in these areas cause higher predictions due to the aforementioned back-transformation function. Thus, one should consider the map of residual kriging variances as well (Figure 5, right). Unfortunately, the values for France are probably not as low as the maps suggest. The data provided by France for the year 2005 were not corrected for a different measurement technique. The measurements are known to be roughly 30% below their official European reference measurement equivalent (MEEDDAT 2008). A time series of interpolated maps is shown in Figure 7 and illustrate the day to day evolution of the interpolated mean PM_{10} concentration levels that may occur.

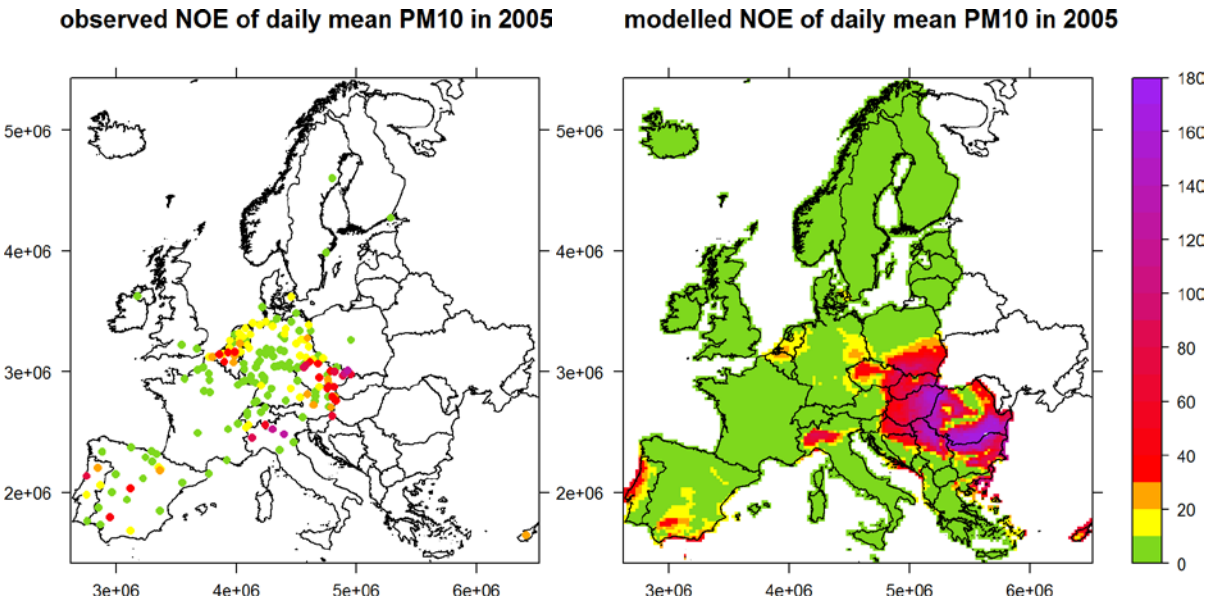


Figure 6: Observed (left) and modelled (right) NOE of daily PM_{10} in 2005.

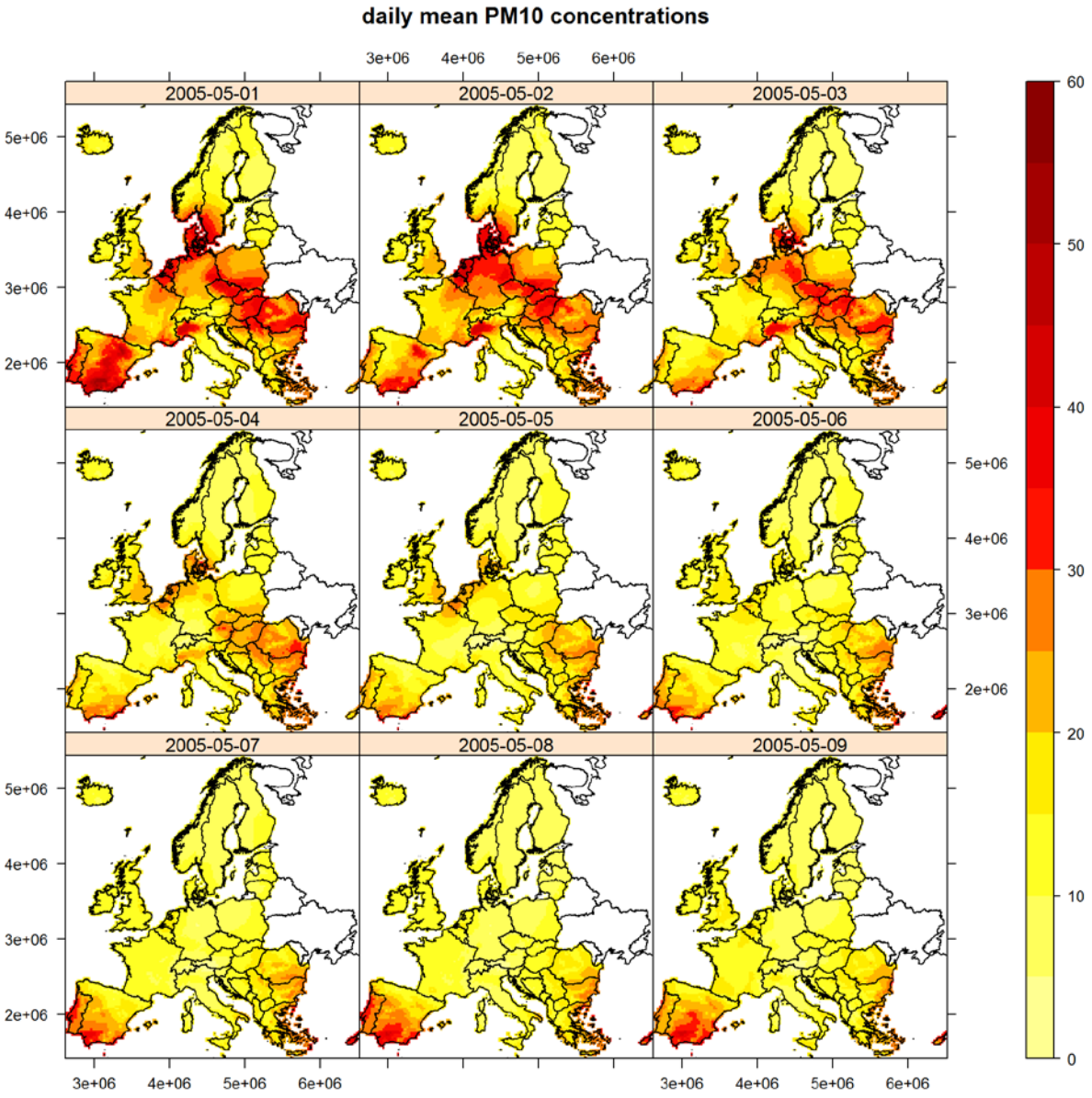


Figure 7: Interpolated maps for daily PM_{10} concentration from May 1 to 9, 2005.

4 Detecting extremes

Investigating the very prominent outliers in the daily and yearly cross-validation plots (Figure 4) reveals a set of stations that typically have several mismatches. Additionally, the mismatches tend to occur at the same day for different stations or temporally and spatially close. Those cases might indeed be extreme observations compared to the overall European daily mean PM_{10} concentrations. However, locations and days showing only a single mismatch might yield candidates for further investigation.

Figure 8 shows that very strong under-predictions typically occur during the summer and winter months. Over-prediction of values seems to be rather rare. Interestingly, a single station is responsible for five of the very strong deviations during winter (red dots). The time series of this station, CZ0TVER, is given in Figure 9. During the summer months, the most extreme values are due to four stations in Portugal (red dots). Their time series for the corresponding months are given in Figure 10. As for any smoothing method, under-prediction of extremes is a property of the kriging method used here as well.

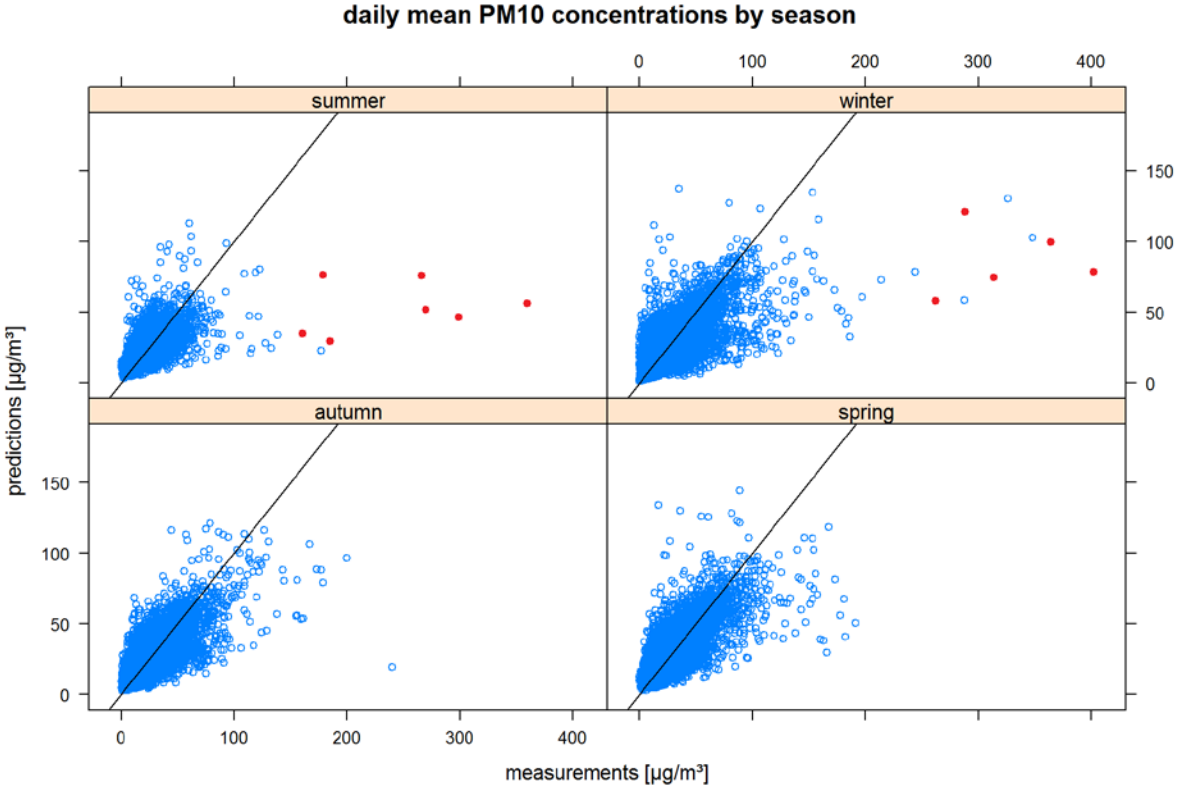


Figure 8: Cross-validation plot of daily mean PM_{10} concentrations by season of 2005.

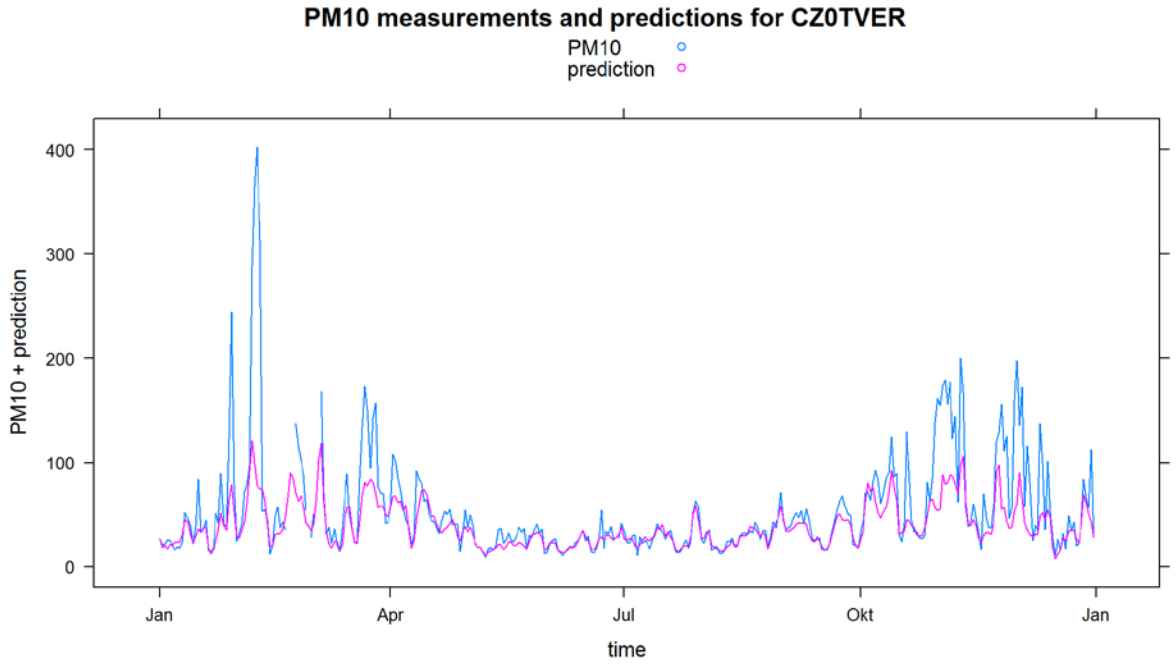


Figure 9: Time series for 2005 of predicted and measured daily mean PM_{10} concentrations for the Czech station CZ0TVER.

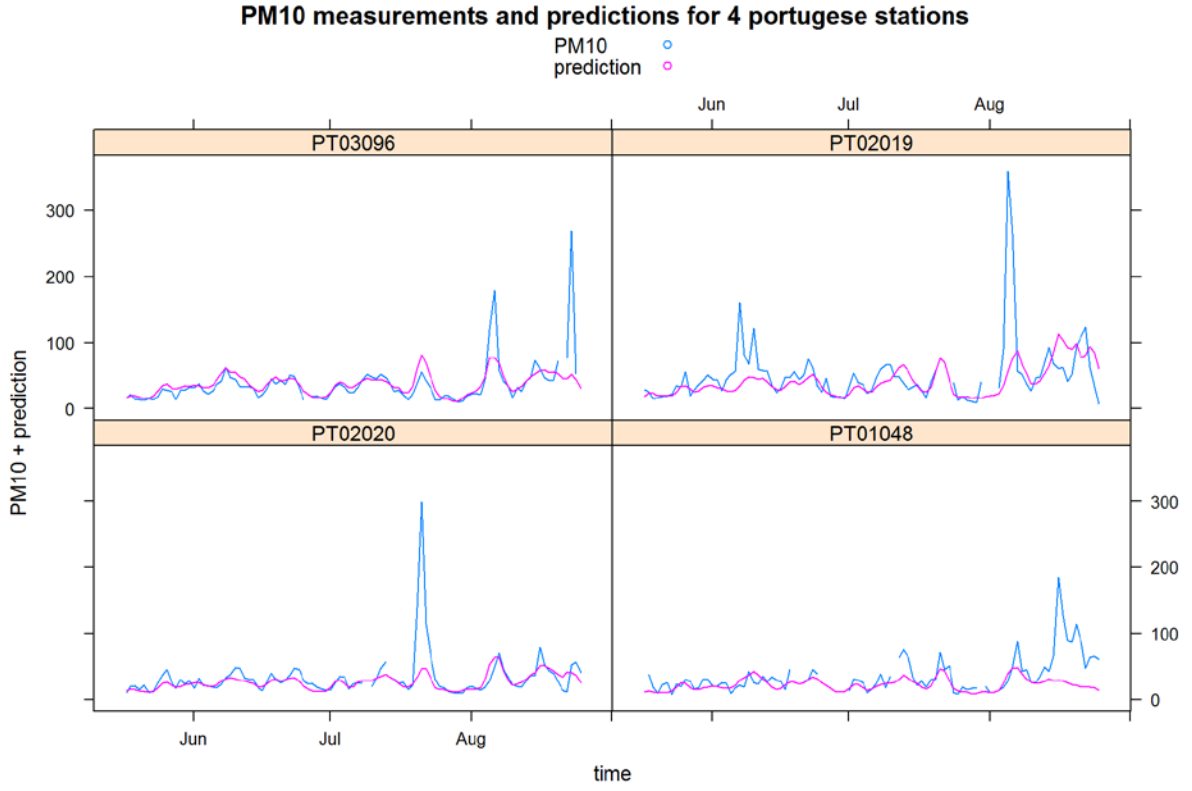


Figure 10: Subsets of four time series for 2005 of stations in Portugal.

5 Discussion

5.1 Is inclusion of spatio-temporal correlation worth the effort?

This report shows that for the spatial interpolation of air quality parameters, it pays off to include temporal correlation to extend spatial-only correlation. This is not surprising, as it could be expected that these correlations would be present, and ignoring (correlated) information is in principle to some extent a loss. Including temporal correlations comes at a cost of a more complex modelling effort, and the statistical gains need to be weighed against the costs of processing.

The comparison of interpolation methods was done based on cross-validation of rural background stations, and was done for one year (2005). The choice of this year was made to allow comparison to earlier method comparison studies (Denby et al., 2008a). Evaluating a method comparison over several years, or from year to year, might yield different results but was out of scope for this report. Thus, further studies are needed to evaluate the robustness of the presented statistical spatio-temporal models that were qualified as best performers in this report based on the 2005 data only. Once deemed robust, these alternative spatio-temporal interpolations or recommendations based on proposed evaluations of asymmetric spatio-temporal covariance models could be applied routinely in mapping and assessment activities that are primarily based on measurement data. The cross-validation study carried out here was done by leaving out, one-by-one, full station time series. As the number of stations is relatively small and the spacing between pairs of stations is relatively large, there is a possibility that this causes another source of variability to the evaluation results. In addition, lack of short-distance station pairs means that cross-validation statistics concern interpolations to distant locations, whereas the use of interpolation for mapping involves to a large extent the interpolation to nearby locations. These two limitations (variability, lack of short distances) are inherent to the data availability and cannot be circumvented without having additional measurement sites.

The symmetric spatio-temporal covariance models evaluated in this study (3D metric (3-f), separable (3-g), product-sum (3-h)) are the most simple methods available. More comprehensive models could for instance address asymmetry in spatio-temporal covariance. In particular for hourly or daily air quality values, it may be expected that the interplay of emissions and weather cause air quality patterns that move over Europe. Seen as a stochastic process, such movement can only be captured by a transport model or by an asymmetric spatio-temporal covariance model. Capturing spatio-temporal dependence by covariances implies that, possibly after transformation, a multivariate Gaussian model is capable of capturing the dependence structure.

Copulas may serve as a useful tool to model asymmetric spatio-temporal data (Gräler et al., 2010). They capture the dependence structures of multivariate distributions independent of the distributions' margins and allow to change continuously over the range of the observations. Thus, pairs of low, high or mixed observations may exhibit a different strength of dependence. Additionally, dependencies may be asymmetric such that tomorrow's measurement may depend differently on today's value than vice versa. The fact that dependencies may change for different magnitudes of the observations allows to model rural and urban concentrations in a single process considerably improving the resolution of the measurement network. Spatial copulas as presented by Gräler et al. (2011) can be used to model spatial and similarly spatio-temporal data and serve as a probabilistic tool to interpolate

spatial and spatio-temporal data. Thus, copulas might also be a useful tool to model PM_{10} concentrations and similar data contained and commonly interpolated in AirBase.

When computing yearly means from daily estimates (Section 3.2, Figure 5), the correct kriging variances were not computed for these mean values. We followed the approximation given by Denby et al. (2008b), but the better, direct approach would be to use block kriging with blocks reflecting a year (i.e., with blocks aggregating over time, as opposed to space). However, before implementing block kriging in the routinely annual mapping exercise research is needed to select the appropriate models and proof its robustness across a series of different years. Given a sufficient robustness, block kriging could be implemented as default method.

5.2 Uncertainty and nugget effects

Measurement uncertainty, the quantitative expression of measurement errors, is hard to obtain from in situ sensors when no simple replicates are taken under controlled circumstances. Every difference between two simultaneous measurements is the result of the combination of measurement error and variability (in space, in time) of the measured PM_{10}.

Gerboles and Reuter (2010) suggest using the spatial variability to inform about measurement errors: they argue that the nugget variance is an upper limit to the measurement error. Figure 11 shows time series of variogram parameters (including nugget variances) of residuals of log transformed PM_{10} for daily variograms over 2005. It shows large variability over time.

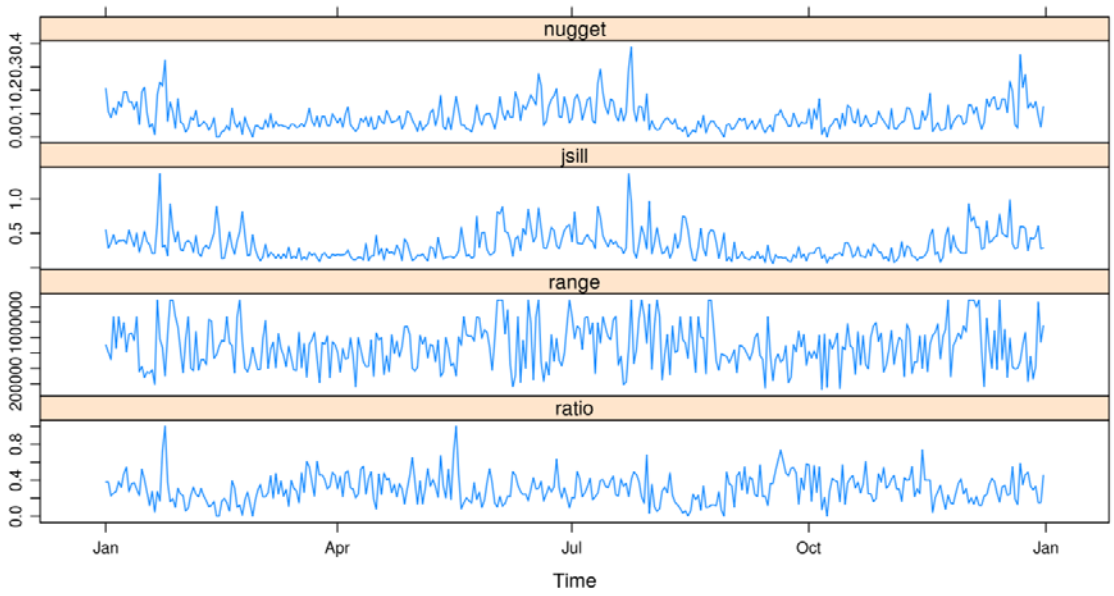


Figure 11: Development of daily variogram parameters of residuals of log transformed daily mean PM_{10} concentrations during 2005. "jsill" indicates the sill (nugget + partial sill, or the variance of the process); "ratio" indicates the ratio of nugget over sill.

The high temporal variability in nugget variances can be attributed to both natural (temporal) variability of the spatial correlation and variability attributable to the problem of fitting a variogram from limited measurement information. The problem of fitting a variogram from measurement information tends to be further augmented by a lack of short spatial distances between observation locations, which is typical for in situ sensor monitoring networks. The assumption here is that nearby monitors yield redundant information, and inter-monitor

distances are rather maximized to obtain a spatially regular distribution against the least monitoring costs.

A potentially more stable procedure as an alternative to using (spatial) nugget variances to estimate an upper limit to the measurement errors would be to use temporal variability, possibly in combination with spatial variability. A first attempt could be to use (half of) the variance of the lag-one differences, in time series of (hourly) measurements series, as a potentially much smaller upper limit to measurement error. This could also reveal station-specific difference that could be worth investigating, or location-specific temporal changes in (upper limits to) measurement errors. A better and more comprehensive attempt would be to use spatio-temporal hierarchical modelling (Cressie and Wikle, 2011) where at the first level the data model essentially isolates the measurement error as an explicit term.

6 Conclusions

In this study we have investigated a set of different approaches to utilize the spatio-temporal dependence when interpolating daily and annual mean PM_{10} concentrations of rural background stations across Europe. Beside rather simple methods, three full spatio-temporal kriging approaches have been presented. A cross-validation has been conducted in order to evaluate the different interpolation methods. For comparability to the interpolation approaches in Denby et al. (2008a), all methods are applied to observational data of the year 2005.

We have shown how a daily, instead of an annual interpolation method, improves the prediction of yearly mean PM_{10} concentrations. The MAE and the RMSE decreased by 7% (see Table 3: improvement of "3D [100] (3-f)" relative to the single annual interpolation using residuals of log-transformed annual mean concentrations "yrl. means") and the adjusted R^2 of the linear model for predictions and measurements increased by 12%. We found that the 3D kriging approach (3-f, with 100 neighbours) provided the best statistical results for almost all statistical indicators for the 2005 data assessed. In particular it was found to improve the prediction of the number of days exceeding the limit of 50 µg/m³ by reducing the RMSE for NOE by 10% and the MAE by 5% compared to the simpler approaches 3-a to 3-e (see Table 3). However, these improvements are relatively modest. It is likely that the exploited temporal dependencies are responsible for this improvement but further studies are needed to evaluate the robustness of the presented procedures for a number of different years.

Furthermore, asymmetric dependence structures, as they are likely to be found in spatio-temporal data where transport in a particular direction takes place, are not captured by the simple symmetric methods evaluated in this report. More flexible methods for modelling spatio-temporal dependence that can deal with asymmetries, for instance spatio-temporal copulas (Gräler et al., 2010) may be used to better reflect spatio-temporal variability in (observational) air quality variables. Beside additional evaluation of the robustness the 3D kriging approach for a series of years, we recommend to investigate in addition the potential of the copulas approach in further studies. However, before implementing one of such spatio-temporal interpolation methods as improved default in the routinely applied mapping exercise, they should have proven their robustness over several data years and for a set of statistical indicators.

Beside the improvement of the statistical measures, daily interpolations provide additional advantages. As briefly demonstrated in Chapter 5, daily predictions allow for a spatio-temporal assessment of outliers incorporating model results. Furthermore, mismatching stations or seasons can be identified. Routinely checking measurements against their predictions will detect deviations from the model. The performance of this outlier detection method has to be compared with other approaches integrating spatial information from several time series. Outliers identified this way by no means need to indicate errors; they indicate merely extremes in the context of the statistical models entertained. Nevertheless, identifying such extremes may be a useful first step leading to careful re-examination of particular sites. Several promising test methods identifying 'questionable' extremes are provided in Gerharz et al. (2011). Looking into NOE of daily mean PM_{10} concentrations, one can identify and study the spatio-temporal distribution of the exceedances using daily predictions in a more natural way. It is questionable whether the interpolation of the yearly number of exceedance days registered at every measurement station is a sensible measure. The temporal aggregation of exceedance days to a single yearly number neglects the temporal dependence and distorts the spatial dependence structure.

A full uncertainty investigation needs to take the uncertainty associated with the multiple linear regression into account as well. Here, we only considered the kriging variance, which

reflects the configuration of measurement locations. Beside the inclusion of measurement stations reporting less than 75% of their data coverage potential for a whole year, no structural differences could be detected on the log-scale. Further probabilistic measures should be evaluated to communicate uncertainties within the policy field. The advances made so far are not yet sufficient in this regard. One way of improving the uncertainty measure would be to encapsulate the full interpolation procedure in a Monte Carlo sampling routine. Thus, multiple predictions would be generated and the variability of these realisations would provide a measure of uncertainty. As outlined in the last paragraph of section 5.1, block kriging would improve the uncertainty measure provided for the yearly means. In order to reduce the sub-grid variability, we recommend to further investigate non-stationary covariance models and anisotropic kriging approaches. The methods developed and discussed in this report can easily be applied to other measures of air quality as well. However, the findings in terms of the best method might be different. The results are strongly related to the spatio-temporal dependence structure of the phenomenon and their robustness should be evaluated for multiple years, preferably in a combined capture of asymmetry in the expected dependence structures. Once deemed acceptable, alternative spatio-temporal interpolation methods could be applied routinely in future mapping and assessment activities that are based on measurement data primarily.

References

AirBase, European air quality database, version 5. EEA. http://airbase.eionet.europa.eu/

Cressie, N., C. Wikle (2011). Statistics for Spatio-Temporal Data. Wiley, New York.

De Iaco, S. (2001). Space–time analysis using a general product–sum model. Statistics Probability Letters, 52(1), pp. 21 - 28

Denby, B., J.Horálek, P. de Smet, F. de Leeuw and P.Kurfürst (2008a). European scale exceedance mapping for PM_{10} and ozone based on daily interpolation fields. ETC/ACC Technical Paper 2007/8. http://acm.eionet.europa.eu/reports/ETCACC_TP_2007_8_spatAQmaps_dly_interpol

Denby B., M. Schaap, A. Segers, P. Builtjes and J. Horálek (2008b). Comparison of two data assimilation methods for assessing PM_{10} exceedances on the European scale. Atmospheric Environment. 42, 7122-7134.

Gerboles, M., H. Reuter (2010).Estimation of the Measurement Uncertainty of Ambient Air Pollution Datasets Using Geostatistical Analysis. Publications Office of the European Union. http://publications.jrc.ec.europa.eu/repository/handle/111111111/15035

Gerharz L, B Gräler, E Pebesma (2011). Measurement artefacts and inhomogeneity detection. ETC/ACM Technical Paper 2011/8. http://acm.eionet.europa.eu/reports/ETCACM_TP_2011_8_artefacts_inhom_detection

Gräler, B., H. Kazianka & G. M. de Espindola (2010). Copulas, a novel approach to model spatial and spatio-temporal dependence. In K. Henneböhl, L. Vinhas, E. Pebesma, & G. Cãmara (Eds.), GIScience for Environmental Change Symposium Proceedings, ifgiprints,40: 49 - 54, AKA Verlag.

Gräler, B. & E. Pebesma (2011): The pair-copula construction for spatial data: a new approach to model spatial dependency. Procedia Environmental Sciences, 7, pp. 206 - 211, Elsevier.

Horálek, J., B. Denby, P. de Smet, F. de Leeuw, P. Kurfürst, R. Swart and T. van Noije, (2007). Spatial mapping of air quality for European scale assessment. ETC/ACC Technical paper 2006/6. http://air-climate.eionet.europa.eu/reports/ETCACC_TechnPaper_2006_6_Spat_AQ

Pebesma, Edzer (2011). spacetime: classes and methods for spatio-temporal data. R packageversion 0.5-2. http://CRAN.R-project.org/package=spacetime

MEEDDAT (2008). Bilan de la qualite de l'air en France en 2007. Ministère de l'écologie, de l'énergie, du développement durable et de l'Amenagement du territoire Paris

Pebesma, Edzer (2004). Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30: 683-691.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.r-project.org